

Seguiment d'una pilota de bàsquet: vers la detecció automàtica d'infraccions

Adrià Arbués Sangüesa, Universitat Pompeu Fabra, arbues6@gmail.com

Abstract— Aquest projecte es basa en la creació d'un *software* que té per objectiu detectar, de forma automàtica, dues infraccions típiques en els partits de bàsquet professional: els tirs fora de possessió i els taps il·legals. La part més important del projecte és la creació d'un algorisme que, donada una seqüència de vídeo, duu a terme un seguiment precís de la pilota. Amb l'esfèrica localitzada, s'empren dues tècniques diferents per a la detecció de les infraccions esmentades: codificació de color per als taps il·legals i extracció de fotogrames en el cas dels tirs fora de temps. Aprofitant la tecnologia ja existent, la implementació d'aquesta eina no suposaria cap cost afegit i seria de gran ajuda per als àrbitres.

Paraules Clau— seguiment de vídeo, tecnologia aplicada a l'esport, bàsquet, OpenCV, *Instant Replay*.

I. INTRODUCCIÓ

En els darrers anys, l'increment d'ús de noves tecnologies a l'esport ha estat una constant. Més concretament en el món del bàsquet, una eina tecnològica ha regnat sobre la resta: l'*Instant Replay*. Aquest és un *software* amb el qual els àrbitres poden consultar en temps real, a través d'una pantalla situada a la taula d'anotadors, la repetició d'una jugada conflictiva per tal donar un veredict encertat. Ara bé, l'*Instant Replay* encara no és una eina perfecta, ja que els àrbitres tan sols poden veure iterativament la mateixa jugada des de determinats punts de vista, que potser són insuficients. D'altra banda, també s'ha de tenir en compte la pressió que reben els àrbitres quan revisen una jugada: tot el públic està pendent de la decisió que ells prenguin, i si aquesta no afavoreix el seu equip, els escridassaran (tinguin raó o no); aquest és un factor afegit que pot influir en el criteri arbitral.

Donat aquest context i vist que les meves dues passions són precisament bàsquet i tecnologia, l'objectiu d'aquest projecte és crear des de zero una aplicació que pugui millorar l'ús actual de l'*Instant Replay*, un projecte ambiciós donada la nul·la recerca en aquest camp d'estudi. Cal destacar que l'objectiu no és substituir l'*Instant Replay* sinó millorar l'eina, mitjançant l'obtenció de veredictes automàtics (i alhora precisos) a partir del post-processat d'una seqüència de vídeo.

Com que detectar totes les infraccions existents era impossible, s'havia de definir l'abast del projecte tot buscant infraccions que involucressin un tipus de processament de vídeo similar. Les dues escollides van ser les següents:

- Taps il·legals: quan un jugador de l'equip ofensiu efectua un llançament a cistella, els jugadors defensors poden tocar la pilota sempre i quan aquesta es trobi en trajectòria ascendent. En cas contrari, la interferència és considerada il·legal i, per

tant, és una infracció que penalitza l'equip defensor. Per tal de detectar els taps il·legals, n'hi hauria prou amb crear una finestra independent; en aquesta, s'hi pintarien punts verds i vermells sobre píxels pels quals ha passat la pilota. Aquesta codificació de color serviria per veure si la pilota es troba en trajectòria ascendent (punts verds) o descendent (vermells).

- Tirs fora de possessió: en els partits de bàsquet, hi ha dos criteris temporals a tenir en compte: la possessió (temps de què un equip disposa per efectuar un llançament a cistella) i el temps de partit. La norma dicta que, quan un d'aquests dos temps és igual a zero, qualsevol acció serà invalidada sempre i quan la pilota encara estigui en contacte amb el jugador.

Aprofitant que quan el temps/la possessió s'acaba s'encenen uns leds vermells en el taulell de la cistella, la detecció d'aquesta infracció podria ser l'extracció de la posició de la pilota en el *frame* exacte en el qual aquests llums s'engeguen. Quan s'encenen els llums, sona a més una botzina perquè l'àrbitre tingui també un estímul sonor, així que l'extracció d'aquests fotogrames serà considerada "sobre la botzina".

En aquest article, els materials i la metodologia emprada es resumeixen a la Secció II, els resultats i el balanç d'avantatges i desavantatges s'expliquen a la Secció III i, finalment, se n'extreuen conclusions a la Secció IV.¹

II. METODOLOGIA

A. Materials Utilitzats (programació i enregistrament)

Per tal de començar el projecte, calia decidir el llenguatge de programació a emprar i trobar el material audiovisual adient. D'una banda, es va optar per fer servir la llibreria *Open-Source* de visió per computador de C++ anomenada *OpenCV* [3]. D'altra banda, totes les seqüències es van registrar a les instal·lacions del Club Bàsquet Vilassar de Mar amb càmeres *GoPro Silver Hero 3+* (un total de 3 càmeres gravant a 60 fotogrames per segon), ja que la majoria de materials referents a seqüències de partits professionals tenien llicències que podien originar problemes.

B. Detecció de Taps Il·legals

Com ja s'ha esmentat, la detecció de taps il·legals serà possible després d'haver dissenyat un algorisme robust que detecti i segueixi la pilota. La idea inicial d'aquest algorisme era aplicar un detector de siluetes i llavors implementar la *transformada de Hough* [1] per trobar aquells objectes que

¹ Noti's que aquest article és tan sols una breu mostra utilitzada per la XVII edició de l'Exporecerca Jove, el treball sencer pot ser demanat per correu.

tenen forma circular en el pla 2D. No obstant, donada la velocitat de la pilota en els llançaments, aquesta perdria la condició d'esfèrica (*motion blur*).

Així doncs, es va optar per un mètode més car en termes de computació, però alhora més fiable, basat en l'eliminació de candidats. El primer pas va ser una extracció del fons de la seqüència, per tal d'eliminar tots aquells píxels estàtics que de ben segur no es corresponien a la pilota, ja que aquesta està sempre en moviment; el mètode implementat modela cada píxel com una mixtura de gaussianes (de Stauffer i Grimson [7], amb la corresponent millora de Zivkovic [8]). Un cop detectats els objectes en moviment, calia segmentar-los mitjançant punts i fer-ne un seguiment. Per tal de dur-ho a terme, es va utilitzar el *detector de cantonades* de Harris [5] i es va aplicar l'algorisme de flux òptic dissenyat per Lucas i Kanade [2] (més concretament, la implementació piramidal de Bouguet, que té com a millora un seguiment precís de moviments llargs en un interval de temps curt). En un principi, es clicava sobre l'objecte a seguir, es trobava la *cantonada* de Harris més propera, i se'n feia el seguiment, però aquesta era una idea inviable, tant per les imprecisions que comportava com per la impossibilitat d'interacció dels arbitres en un partit real.

La solució va ser detectar automàticament totes les *cantonades* del primer pla de la seqüència i aplicar un filtre de color, de manera que es descartaven tots els objectes que no fossin la pilota. No era una tasca trivial, ja que l'esfèrica era d'un color molt semblant als objectes del fons i a la pell dels jugadors/es. Així doncs, es va optar per utilitzar una pilota de color blau; una solució que no serviria per un sistema professional però sí per un projecte acadèmic. El filtre estaria basat en valors de l'espai de color HSV (*Hue-Saturation-Value*) i els rangs d'aquest es delimitarien mitjançant l'extracció manual de valors dels tres canals en 3 seqüències de vídeo diferents en 30 fotogrames diferents (90 mostres en total per cada canal). Tal i com s'esperava, els resultats del canal de la tonalitat (*Hue*) van ser molt similars en la majoria dels casos, així que el filtre podia tenir franges molt selectives; d'altra banda, els altres dos canals van ser utilitzats per eliminar aquells blaus molt foscos o clars. Així doncs, els valors finals del filtre van ser: {112-119} *Hue* (rang entre 0 i 180 graus), {>75} *Saturation* i {>80} *Value* (valors entre 0 i 255). Aquells punts trobats mitjançant el *detector de cantonades* de Harris que complien aquesta composició de color eren considerats com a pilota.

Arribats a aquest punt, sorgia un altre problema: si bé en el primer fotograma de la seqüència de vídeo es detectava correctament la pilota (sempre i quan aquesta estigués dins del pla de l'escena), a mesura que se'n feia el seguiment el punt detectat es perdia o es confonia amb el jugador, ja que l'algorisme de seguiment no té en compte el color. Per solucionar aquest impediment, es va implementar un sistema que tenia per objectiu refrescar la posició de la pilota cada 10 fotogrames, determinant-ne la nova posició i fent-ne el seguiment a partir d'aquesta.

L'únic pas que quedava per a la detecció de taps il·legals era el mapeig de posicions on es detectava la pilota i

l'aplicació d'una codificació de color. En un principi, es va optar per la creació d'una finestra independent de la mateixa mida que la seqüència original, per tal de pintar amb punts de color negre els píxels on s'havia trobat la pilota i unir aquells que fossin consecutius temporalment. Abans d'aplicar la codificació de color, però, calia eliminar els *outliers* (punts que, de ben segur, no es corresponien a la pilota) que esdevenien quan dos o més punts complien els requisits del filtre de color. La solució va ser aplicar un criteri de proximitat: el punt més proper al trobat al fotograma anterior seria el considerat com a pilota.

Finalment, es podria aplicar la codificació per color: la posició de la pilota seria pintada de color verd sempre i quan l'ordenada del punt fos major que la de l'anterior fotograma enregistrat; en cas contrari, el punt seria pintat de color vermell. Així doncs, donada la seqüència de vídeo original, s'acaba obtenint la trajectòria de la pilota en una finestra en blanc, on es veuen mitjançant punts verds els intervals en els quals l'esfèrica puja (tap legal, color verd) i vermells en els quals baixa (tap il·legal, color vermell). El procediment gràfic per tal de detectar taps il·legals és mostrat a la Figura 1 al final de l'article.

C. Extracció de fotogrames sobre la botzina

Aprofitant l'algorisme de seguiment de la pilota, la idea inicial per saber si un tir estava dins o fora de temps era fer el mateix mapeig en una finestra independent i pintar de color verd els punts en els quals el temps encara no s'havia acabat i de vermell el cas contrari. Tanmateix, aquest enfocament no suposaria una ajuda per als àrbitres, ja que el que volen saber aquests és si la pilota està en contacte amb el jugador en el moment en què s'encenen els *leds* del taulell. Així doncs, es va canviar d'estratègia.

Donat aquest nou escenari, l'objectiu era detectar la regió del taulell i determinar amb total precisió el moment en el qual s'encenen els llums. Per tal de dur-ho a terme, el programa rebia com a *input* les seqüències de vídeo des de tres punts de vista (amb una de les tres càmeres captant el taulell sencer) i una imatge de mostra del taulell. El primer pas per determinar la regió d'interès va ser detectar els descriptors *SIFT* (*Scale Invariant Feature Transform*) [6] en el primer fotograma, tant de la seqüència des d'on es veia el taulell com de la imatge de mostra. Just després, els descriptors d'ambdues imatges serien emparellats amb un *Brute-Force Matcher*, el qual utilitza un criteri de distància entre tots els possibles punts. Finalment, amb l'algorisme de *RANSAC* (*RANdom SAMple Consensus*) [4], s'estimaria l'homografia que relaciona les dues imatges, i s'identificarien quins píxels corresponen a la regió del taulell en la seqüència de vídeo original.

Sabent la regió on es trobava el taulell, es va crear un filtre de color com l'utilitzat en el cas dels taps il·legals, però acceptant els colors de tonalitat vermella. Es va crear una màscara binària per aquesta mateixa regió i, per cada nou fotograma, es comptava el nombre de píxels vermells (suma de la màscara); el primer fotograma en el qual aquesta suma suposava més d'un 2.5 % dels píxels totals de la regió era el corresponent al primer *frame* amb els *leds* del taulell encesos.

Cal notar que aquest percentatge és un llinard totalment experimental ja que, en les seqüències utilitzades, els llums vermells s'han simulat amb línies vermelles introduïdes amb *Illustrator*, donada la manca de la tecnologia real.

Tenint les seqüències de vídeo sincronitzades des de tres punts de vista diferents, l'extracció de fotogrames sobre la botzina es basaria en la creació d'una variable binària, que indicaria que els *leds* del taulell s'han encès; és en aquest precís instant quan totes les càmeres mostrarien el fotograma corresponent. El procediment explicat és representat gràficament a la Figura 2 al final de l'article.

III. RESULTATS I DISCUSSIÓ

Els resultats obtinguts es poden observar a les Figures 1 (g) i 2 (d, e i f). Tanmateix, en tractar-se d'un projecte de processament de vídeo, les seqüències amb el seu corresponent processat han estat penjades en un canal de *Youtube*, al qual es pot accedir mitjançant el següent codi QR:



A. Avantatges

- Els resultats obtinguts són automàtics i precisos. No involucren cap interacció de l'equip arbitral i en cap cas poden fer nosa.
- El *software* creat no suposa cap mena de cost, donada la quantitat de càmeres que s'utilitzen avui dia per les retransmissions de partits.
- Es tracta d'una eina útil per als àrbitres (*ergo* útil per l'esport), ja que escurçaria el temps emprat per la consulta de l'*Instant Replay*; per tant, el partit en sí tindria menys interrupcions (o, si més no, més curtes).

B. Desavantatges / Possibles millores

- El sistema hauria de ser independent del color. Com ja s'ha esmentat, la pilota utilitzada en els vídeos és de color blau, cosa que facilita la implementació d'un filtre de color ideal. A més a més, no es pot privar un equip de vestir de color blau tan sols per poder aplicar aquest programa.
- Els problemes d'escala i inclinació amb les imatges de mostra quan es vol trobar la regió del taulell poden tenir conseqüències dràstiques. Donada la baixa resolució amb què es feia el processat, la regió del taulell tenia artefactes no negligibles que van fer que el *software* no funcionés amb la majoria d'imatges de mostra.
- Les proves realitzades amb el programa han estat amb seqüències on manca realisme, amb tan sols 2 persones a l'escena i un fons estàtic. En un partit professional hi ha fins a 10 jugadors i 3 àrbitres a la pista i de ben segur que el fons estarà contínuament en moviment.

IV. CONCLUSIONS I *FUTURE WORK*

La hipòtesi plantejada a l'inici del projecte ha estat complerta, ja que s'ha dissenyat un programa complementari a l'*Instant Replay*. Aquest *software* és capaç de detectar automàticament taps il·legals, mitjançant una codificació de color, i tirs fora de possessió, a través del recompte de píxels vermells a la regió del taulell.

Els resultats obtinguts no han estat ni de bon tros dolents, però per tal de millorar aquest projecte, es podria fer una prova real, amb càmeres de qualitat sincronitzades mitjançant fibra, 10 jugadors a pista (i espectadors a les grades) i més punts de vista. Una altra possible millora seria la inclusió d'un percentatge de confiança, per saber com és de fiable el veredictes d'una acció després que l'algorisme hagi fet el seu post-processat. Per tal d'eliminar el filtre de color, es podria fer ús de càmeres tèrmiques, però aquestes encara són una tecnologia emergent.

Amb aquestes millores i creant una *GUI* (*Graphical User Interface*) que funcionés a temps real, el programa estaria llest per ser aplicat als partits de bàsquet professional i tindria un impacte positiu immediat sense cap cost d'implementació.

V. BIBLIOGRAFIA

- [1] Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2), 111-122.
- [2] Jean-yves Bouguet. Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. In Practice, 1(2):1-9, 1999.
- [3] G. Bradski. OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.
- [4] Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- [5] Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, p. 50).
- [6] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- [7] Chris Stauffer and W E L Grimson. Adaptive background mixture models for real-time tracking. Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Cat No PR00149, 2(c):246-252, 1999.
- [8] Zivkovic, Z. (2004, August). Improved adaptive Gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 2, pp. 28-31). IEEE.

Adrià Arbués Sangüesa (28 de Desembre del 1993)



- Educació Secundària: INS Vilatzara, 2005-2009.
 - Batxillerat: INS Vilatzara (modalitat tecnològica), 2009-2011
 - Educació Superior: Grau en Enginyeria en Sistemes Audiovisuals, Universitat Pompeu Fabra, 2011-2015; Màster en Visió, Gràfics i Sistemes Interactius, Universitat d'Aalborg (Dinamarca), actualitat.
- Correu: arbues6@gmail.com
 LinkedIn: <https://www.linkedin.com/in/adriaarbués>

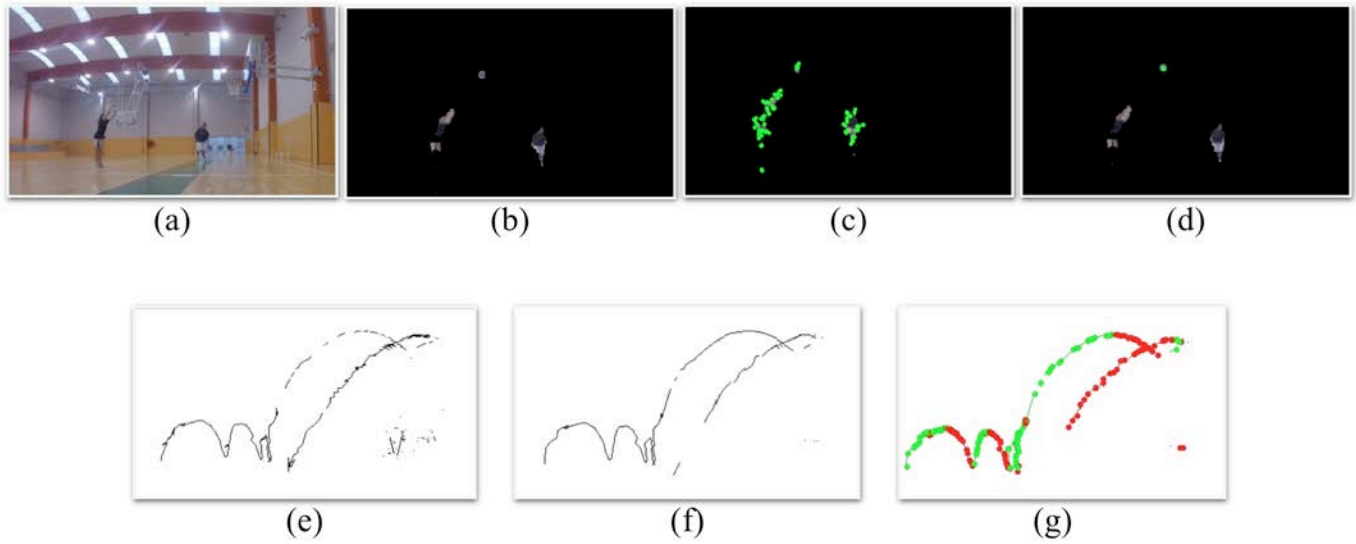


Figura 1: procediment per detectar taps il·legals; (a) Seqüència de vídeo original; (b) Extracció del primer pla de la seqüència; (c) Detecció de totes les cantonades del primer pla; (d) Filtratge per color; (e) Mapeig en 2D dels punts per on passa la pilota; (f) Filtratge d'*outliers*; (g) Codificació de color.

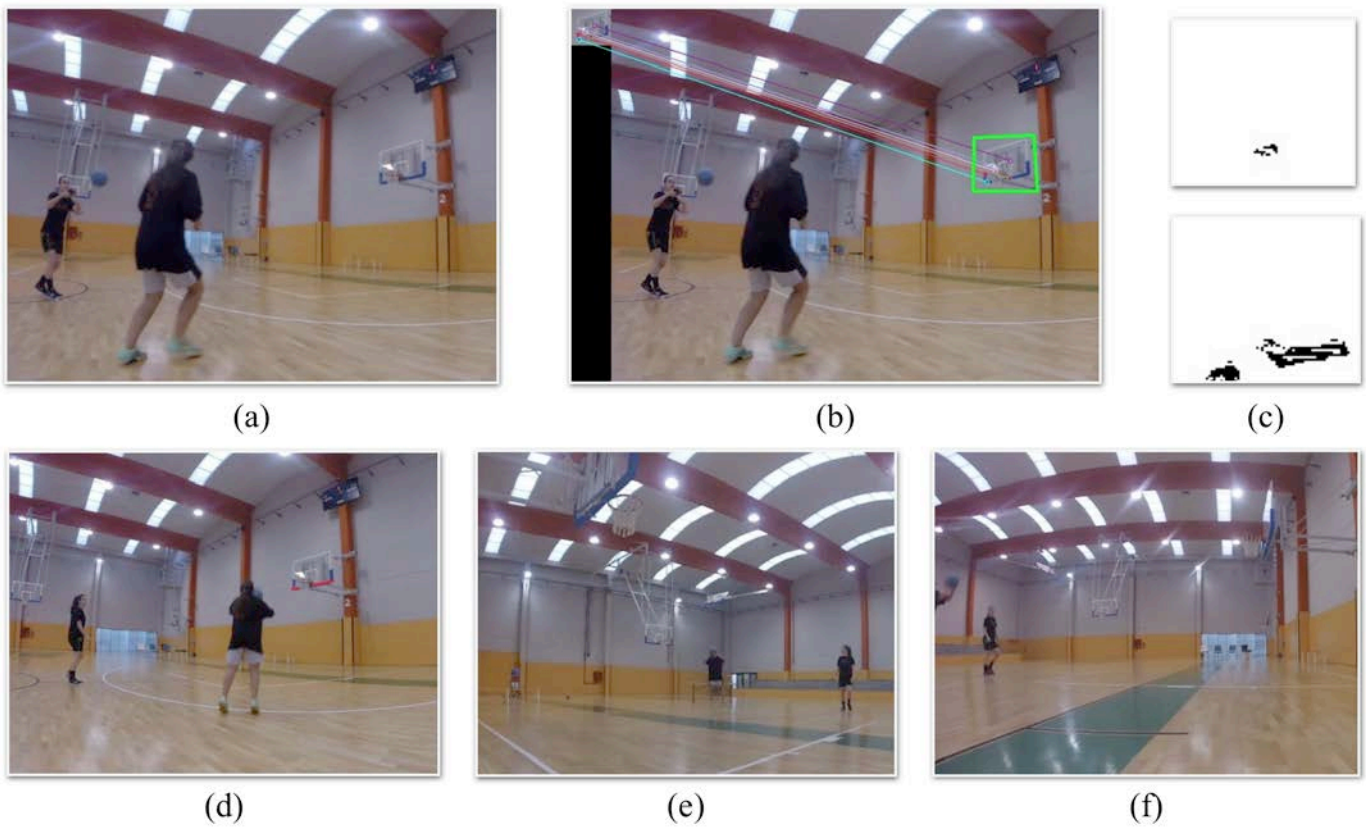


Figura 2: procediment per extreure fotogrames sobre la botzina; (a) Primer fotograma d'una seqüència de vídeo; (b) Emparellament de descriptors entre la imatge de mostra del taulell i el primer fotograma, estimació de l'homografia per trobar la regió del taulell; (c) Màscara binària amb el nombre de píxels vermells dins la regió del taulell: llums apagats (imatge superior) i encesos (inferior); (d, e, f) Fotogrames extrets en el moment que s'encenen els llums des dels tres punts de vista.